Check for updates

the FEBS Journal

REVIEW ARTICLE

DNA structure and function

Andrew Travers[1,2] and Georgi Muskhelishvili[3]

1 MRC Laboratory of Molecular Biology, Cambridge, UK
2 Department of Biochemistry, University of Cambridge, UK
3 Jacobs University Bremen, Germany

Keywords
A-DNA; B-DNA; alternative DNA structures; DNA as an energy store; DNA backbone conformation; DNA elasticity; DNA information; DNA structure; DNA topology; genome organisation

Correspondence
A. Travers, MRC Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge CB2 2QH, UK
Tel: +44 (0)1223 891921
E-mail: at534@cam.ac.uk

(Received 14 November 2014, revised 26 February 2015, accepted 21 April 2015)

doi:10.1111/febs.13307

The proposal of a double-helical structure for DNA over 60 years ago provided an eminently satisfying explanation for the heritability of genetic information. But why is DNA, and not RNA, now the dominant biological information store? We argue that, in addition to its coding function, the ability of DNA, unlike RNA, to adopt a B-DNA structure confers advantages both for information accessibility and for packaging. The information encoded by DNA is both digital – the precise base specifying, for example, amino acid sequences – and analogue. The latter determines the sequence-dependent physicochemical properties of DNA, for example, its stiffness and susceptibility to strand separation. Most importantly, DNA chirality enables the formation of supercoiling under torsional stress. We review recent evidence suggesting that DNA supercoiling, particularly that generated by DNA translocases, is a major driver of gene regulation and patterns of chromosomal gene organization, and in its guise as a promoter of DNA packaging enables DNA to act as an energy store to facilitate the passage of translocating enzymes such as RNA polymerase.

Introduction

DNA is now the predominant genetic material in the living world. But which properties of this long polymer favour this ubiquity? Why DNA and not, say, the structurally very similar double-stranded RNA?

Just over 60 years ago Watson and Crick published their classic paper [1] on the structure of DNA. In it, they emphasized two principal features of the molecule – the complementarity of the base sequences on the two strands and the double-helical nature of the polymer. The base sequence complementarity, with adenine complementary to thymine and guanine complementary to cytosine, provided an elegant molecular explanation for the discovery in the previously decade by Avery, McCarty and Macleod [2] that DNA was likely the 'transforming principle' that enabled the transfer of genetic information between different strains of bacteria. More, it confirmed Chargaff's fundamental discovery of the equivalence of A and T and of G and C bases in double-stranded DNA [3]. Most importantly, the structure implied that the information in the DNA base sequence could possess, by virtue of the complementarity, the ability to be replicated into two identical copies. This insight into the fundamental basis of genetics has underpinned the immense advances in genetic understanding and manipulation over the last 60 years.

Today, however, the feature of DNA that defines the molecule is the fact that the two strands are entwined as a right-handed double helix [4]. DNA is 'the double helix'. Although this double-helical character is not required by the complementarity *per se* – a simple straight ladder structure would fulfil this func-

Abbreviations
DNA, deoxyribonucleic acid; FACT, facilitates chromatin transcription/transactions; FIS, factor for inversion stimulation; HMGB, high mobility protein B; H-NS, histone-like nucleoid structuring protein; RNA, ribonucleic acid.

tion just as well – it does impart crucial physical and chemical properties to the polymer. It is these properties that play a major role in the biological function of DNA. The genetic functions of DNA can thus be understood as the synergism of two properties – a tape containing the information store encoding the sequences of proteins and RNA molecules and a polymer existing as double-helical string enabling the packaging, accessibility and replication of the information store. Crucially, both the coding of proteins and RNA molecules and also the physicochemical properties of the polymer are specified by the base sequence.

DNA as an information store

What is the nature of the genetic information stored in DNA? The distinction between a linear code responsible for specifying the sequences of RNA and protein molecules and also sequence-specific recognition by DNA-binding proteins, and an equally important more continuous structural code, specifying the configuration and dynamics of the polymer extends the informational repertoire of the molecule. Both these DNA information types are intrinsically coupled in the primary sequence organization, but whereas the linear code is, to a first approximation, a direct digital readout [5,6], the structural code is determined not by individual base pairs, but by the additive interactions of successive base steps. The latter code, being locally more continuous, thus has an analogue form [5,6]. Importantly, the manifestation of analogue properties is dependent on the length of the DNA sequence. For example, under physiological conditions, DNA unwinding manifested as melting may be restricted to a short sequence (say up to ~ 10 bp), whereas unwinding in the form of a coiled helical axis may affect hundreds of base pairs [7,8].

Direct and indirect readout of DNA-recognition sites by proteins is a major determinant of binding selectivity. In direct readout, the individual bases in a binding sequence make direct and specific contacts to the protein surface, whereas in indirect readout, the binding affinity depends on recognition of a structure, such as a DNA bend or bubble, whose formation is influenced by DNA sequence, but does not in general require a protein contacting a specific base. In practice, DNA recognition by proteins effectively spans a continuum from completely digital to completely analogue with many proteins utilizing both modes.

For both modes of recognition, the DNA double helix differs from, and is arguably more effective than, the RNA double helix. Direct readout requires intimate contact between exposed chemical groups on both the protein and nucleic acid surfaces. For DNA recognition, direct readout in most examples takes the form of a DNA-binding motif being inserted into the major groove. In this groove, different exocyclic groups of the bases in a pair are exposed compared with those in the minor groove (Fig. 1). Consequently, although A–T and T–A base pairs in a sequence are distinguishable by the position of the thymine methyl group charge pattern in the major groove, in the minor groove, the exposed charge patterns of T–A and A–T base pairs are identical. Similarly, the charge patterns of C–G and G–C base pairs in the major groove are distinguishable by the relative position of the 4-amino group of cytosine. Again, however, there is little difference in the relative spatial arrangements of the charge patterns of C–G and G–C base pairs in the minor groove. The major groove thus provides more

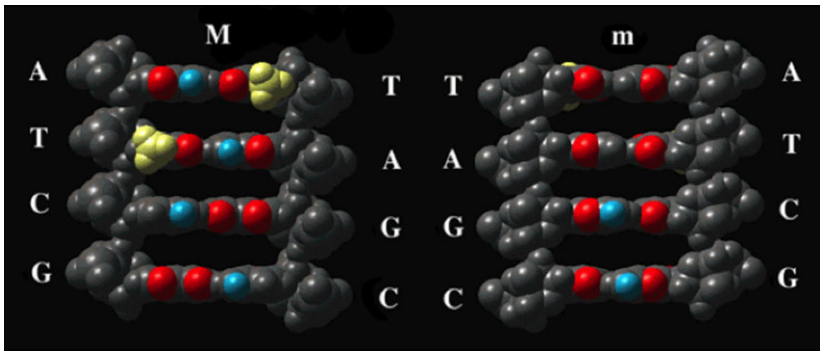

Fig. 1. Exposure of chemical groups of nucleotide bases in the major and minor grooves of DNA. M, major groove; m, minor groove. Yellow, thymine 5-methyl group; blue, basic groups: adenine 6-amino group (major groove), cytosine 4-amino group (major groove) and guanine 2-amino group (minor groove); red, exposed cyclic nitrogen atoms and oxy- groups. Note that the presence of the thymine 5-methyl group in place of hydrogen in uracil the enables A–T base pairs to be distinguished from T–A base pairs in the major groove. (Adapted with permission from IMB Jena Image Library).

17424658, 2015, 12, Downloaded from https://febs.onlinelibrary.wiley.com/doi/10.1111/febs.13307 by National Health Fund in Poland, Wiley Online Library on [01/08/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

sequence information than the minor groove. However, importantly, the wide and shallow morphology of the DNA major groove is in stark contrast to the narrow and deep structure of the RNA major groove. This pattern is reversed for the minor groove. For a protein DNA-binding motif, particularly one containing an α-helix, access to the DNA major groove is more facile than to the minor groove. This fundamental difference between DNA and RNA follows directly from their chemical structures. Whereas DNA can adopt (at least) two forms of right-handed double-helical structures, A-DNA and B-DNA (Fig. 2A), RNA can only form an A-type double helix because of the steric restrictions imposed by the 2′ hydroxyl residue on ribose [9–12]. The B-DNA structure, that proposed by Watson and Crick [1], is most stable at high humidity, but converts to the A-form as the water activity is lowered [13]. On this argument, it is the ability to adopt the B-form that facilitates direct access to DNA sequence information.

Not only does the A → B transition affect direct readout, it also changes the physicochemical properties of the polymer. An A-type double helix is, on average, stiffer than a B-type double helix and consequently distortion of A-DNA to a particular bent configuration is energetically less favourable than for the corresponding distortion in B-DNA [14]. Such differences would be expected to favour B-DNA as the preferred substrate for packaging involving tight DNA bending.

Although the formation of a B-type structure is a crucial aspect of DNA functionality the factors which shift the A ↔ B equilibrium are, apart from water activity, poorly understood. One aspect is base-type. In principle, the coding capacity of DNA can be achieved not only by the canonical A–T and G–C base pairs, but also by other possibilities. For example, a DNA polymer with diaminopurine–thymine (DAP–T) and hypoxanthine–cytosine (H–C) base pairs with, respectively, three and two interbase hydrogen bonds (Fig. 3) would, in principle, present a similar potential for protein recognition and thermal stability [15]. Other variations would be DNA molecules in which all the base pairs contain either two or three hydrogen bonds [16]. However, not only do the component bases specify a digital code, they also affect the physicochemical properties of the molecule. For example, DNA molecules with a reversed pattern of hydrogen bonding (DAP–T and H–C base pairs) more readily adopt an A-type conformation than DNA with the canonical base pairs [16,17]. This is because the properties of the double helix depend not only on the base-pairing capacity of the constituent bases, but also on the stacking interactions between adjacent base pairs. Changing base-pairing interactions by effectively transferring a 2-amino group from guanine to adenine (thereby creating hypoxanthine and diaminopurine) changes the overall stacking because the charged 2-amino group, by being in a different immediate chemical environment, also affects the dipole moments asso-

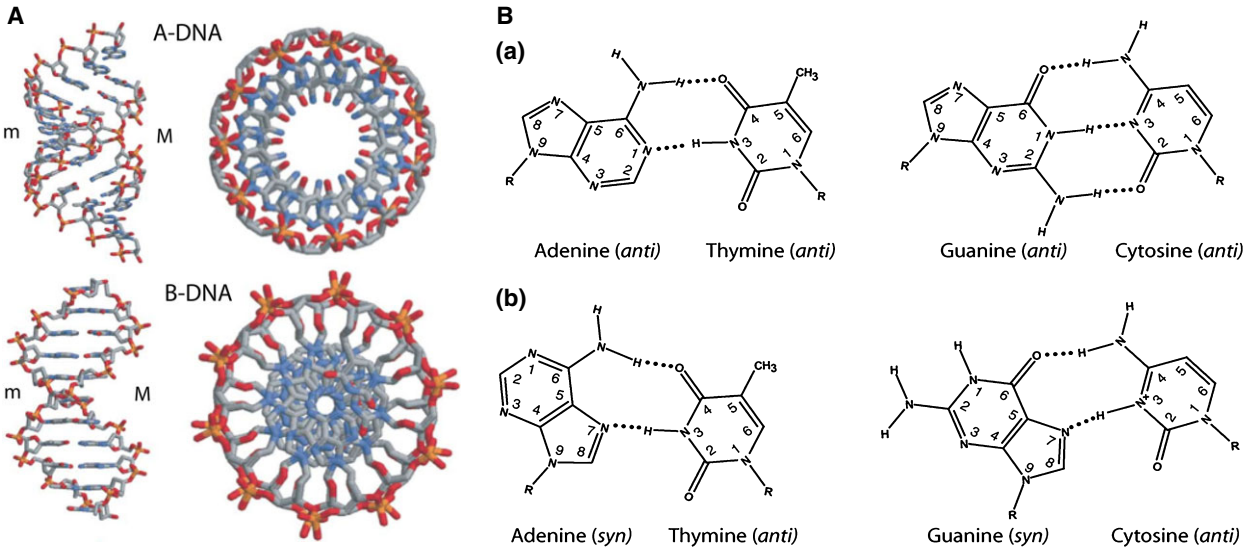

Fig. 2. (A) Structures of A-DNA and B-DNA. Note the difference in groove width and the relative displacements of the base pairs from the central axis. Reproduced with permission from Arnott [12]. (B) A–T and G–C base pairs shown for Watson–Crick pairing (a) and Hoogsteen pairing (b). *syn* and *anti* indicated different sugar conformations. Reproduced with permission from Johnson *et al.* [124].

17424658, 2015, 12, Downloaded from https://febs.onlinelibrary.wiley.com/doi/10.1111/febs.13307 by National Health Fund in Poland, Wiley Online Library on [01/08/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

ciated with individual base pairs and consequently the stacking interactions between base pairs (Fig. 4) [18,19]. In other words, the ability to assume the B-conformation, which confers on DNA an important aspect of its unique genetic role, is itself dependent on base-type and in particular on A–T and G–C base pairs. Although this might constitute a reason for the selection of these base pairs in most DNA molecules no such simple argument can be advanced for the use of A–U and G–C base pairs in RNA, although even in RNA the stability of different base-steps and hence of the double-helix itself is likely dependent on the precise nature of the constituent base pairs.

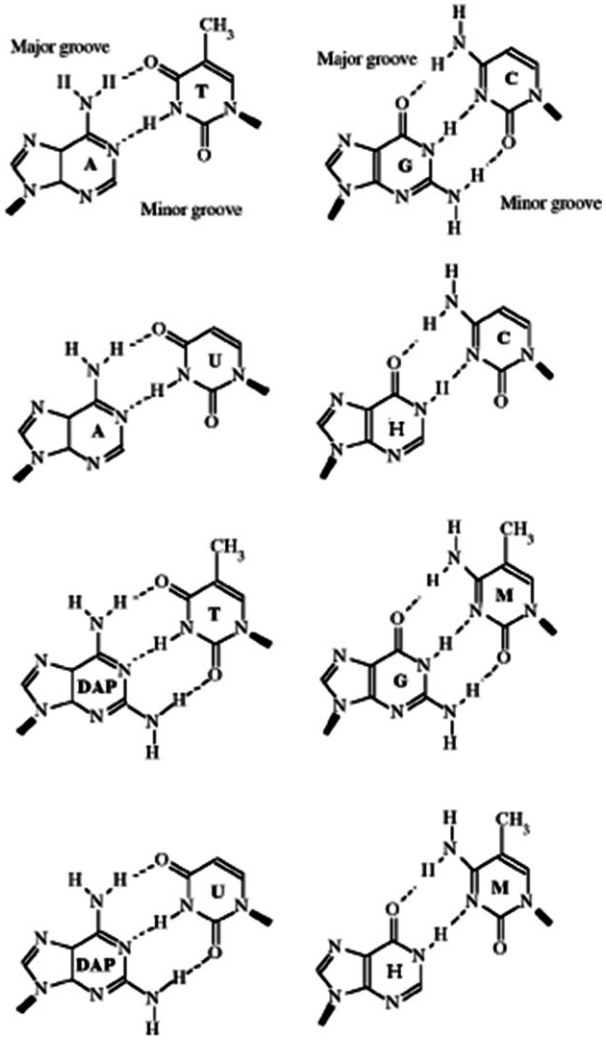

Fig. 3. Structures of alternative base pairs maintaining Watson–Crick hydrogen bonding. A, adenine; C, cytosine; DAP, diaminopurine; G, guanine; H, hypoxanthine; U, uracil. Adapted with permission from Bailly *et al.* [15].

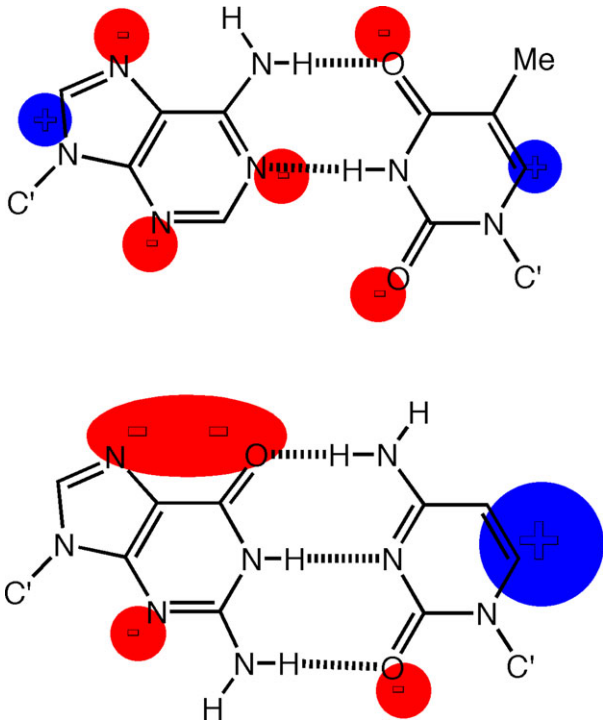

Fig. 4. Dipole moments of A–T and G–C base pairs. Reproduced with permission from Hunter [125].

Alternative modes of sequence recognition

The gold standard of sequence recognition in the double helix employed in both transcription and replication utilizes the base pairing rules formulated by Watson and Crick [1]. However, as they acknowledged, bases can pair in different ways. In particular, a different base-pairing geometry, the Hoogsteen base pairs, in which the purine base is rotated relative to that in the standard base pair, reduces the distance between the C1 carbon atoms of the associated sugar moieties [20] (Fig. 2B). Although this type of base pairing is incompatible with the structure of the canonical DNA double helix, it is found in other structural forms of DNA, notably H-DNA and G-quadruplexes [21–24] (see below). Similarly, yet another type of noncanonical base pair has been postulated to stabilize the *i*-motif formed by sequences complementary to those forming G-quadruplexes [25].

Yet another form of DNA–DNA interaction, which has received relatively scant attention, is the ability of two double helices of the same sequence to align with each other [26,27]. The attractive force causing this DNA self-assembly might function in biological processes such as folding of repetitive DNA, recombination between homologous sequences, and synapsis in

17424658, 2015, 12, Downloaded from https://febs.onlinelibrary.wiley.com/doi/10.1111/febs.13307 by National Health Fund in Poland, Wiley Online Library on [01/08/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

meiosis. But how is this type of homologous recognition effected? A possible mechanism is the mutual alignment of the electrostatic signature of a base sequence [28]. Another, not exclusive, suggestion is that the flipping out of bases from the double helices may also be involved [26].

DNA as a conformationally flexible and dynamic polymer

In genomes, DNA molecules are generally very long, thin polymers with a diameter of 2 nm and a length that can extend to 10^8–10^9 nm. As an information store, not only must DNA be able to encode the genetic information required to specify proteins, but also it should be packaged in a compact form that allows the accessibility of that information to be regulated. In turn, the functional accessing of information may also involve structural changes in the double helix itself. However, the very nature of DNA – again an immensely long, very thin polymer – requires that within the cell the molecule be compacted into a small volume while maintaining accessibility. These requirements for compaction, accessibility and structural modulation imply that DNA be both flexible and able to change conformation in response to enzymatic manipulation.

The DNA molecule may be modelled as an extremely long thin string of moderate elasticity that can be bent into the configurations required for packaging. Both the preferred direction of bending and the stiffness are sequence dependent [29–31]. A directional bending preference, or bending anisotropy, facilitates the wrapping of DNA on a complementary protein surface. However, such a preference also implies that anisotropy increases the overall stiffness because it reduces the degrees of bending freedom. In other words, by reducing bending freedom, the intrinsic bending entropy of the sequence is reduced and on binding to a preferred protein surface there is a corresponding reduction in the entropic penalty [32]. Such bending preferences are important determinants of the binding of both enzymatic manipulators of DNA and the abundant, so-called 'architectural', DNA-binding proteins [33], which direct the local packaging of the polymer. The archetypical example of this mode of packaging is the nucleosome core particle – the fundamental unit of DNA packaging in eukaryotic chromosomes – in which 145 bp of DNA are wrapped in 1.6 turns tightly around a histone octamer [34]. The signature of bending directionality is the presence of alternating short stretches of G/C-rich and A/T-rich DNA sequences in the helical phase [31,35,36]. Such an organization confers bending anisotropy because G/C and A/T-rich sequences favour, respectively, wide and narrow minor grooves [31]. Consequently, because in tightly bent DNA, both DNA grooves are narrowed on the inside of a bend and widened on the outside, G/C-rich sequences are favoured in which the minor groove points outward and A/T-rich sequences in which the minor groove points inward.

Even when free in solution, certain DNA sequences can confer a preferred axial configuration [29,37–39]. Such sequences contain base-steps that are conformationally rigid [18,40]; that is, the base-step can adopt only a limited range of conformations. The most notable of these sequences are stretches of oligo(dA)(dT) in which the AA/TT base-steps are stabilized by bifurcated hydrogen bonds [41–43]. By themselves, these sequences are straight, but when juxtaposed with G/C-rich sequences, the whole sequence adopts a curved configuration [44]. Such intrinsically curved molecules can, when mimicking the curvature of DNA on the histone octamer, facilitate nucleosome formation [45], whereas long straight stretches of oligo(dA)(dT) have a lower affinity for the octamer and can serve to phase nucleosomes *in vivo* [45].

Not only must DNA be bendable in order to be packaged efficiently, but the copying of the DNA sequence during transcription, or DNA replication, requires the separation of the two strands of the double helix, a transition in which the double helix is untwisted to form a bubble. The initiation of copying at the points at which strand separation is nucleated is facilitated by highly localized less stable DNA sequences with lower stacking and melting energies [46]. Of these, the least thermally stable base-step is TpA [47,48]. Such localized untwisting of DNA affects both DNA melting and also DNA bending. When a bubble is formed in the DNA double helix, not only is its bending flexibility increased [49], but so too is the directionality of sequence-directed bending becomes attenuated and more isotropic.

DNA as an energy store

An often overlooked function of DNA in a cell nucleus or bacterial nucleoid is that it can act as an energy store for facilitating the transit of DNA and RNA polymerases. This emergent property is a direct consequence of the double-helical character of the molecule. Not only does it exist as a simple intramolecular interwound coil, but under torsional stress, the DNA chain can adopt a coiled configuration, or supercoil [50,51]. Such supercoils have a higher intrinsic energy than DNA molecules not subject to torsional stress. Within both the eukaryotic nucleus and the bacterial nucleoid supercoiling is ubiquitous [52–54].

17424658, 2015, 12, Downloaded from https://febs.onlinelibrary.wiley.com/doi/10.1111/febs.13307 by National Health Fund in Poland, Wiley Online Library on [01/08/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

An open circle represents a state in which the DNA molecule, under the prevailing environmental conditions, occupies an energetic minimum. It is 'relaxed'. However, enzymatic manipulation of DNA using energy from ATP can alter the torsional state of DNA inducing more coiling – 'supercoiling' – within such a closed system. This coiling can be 'positive' – overwinding – in the same sense as the DNA double helix or 'negative' – underwinding – in the opposite sense [51] (Fig. 5). Enzymatic manipulation of DNA supercoiling can take two forms. On the one hand, enzyme, termed a 'topoisomerase', can bind to a single site and directly change the coiling [55] and on the other hand, coiling can be changed by the movement under particular constraints of a protein, or protein complex, such as RNA polymerase along the DNA [56–58]. An example of the first case is DNA gyrase, a bacterial topoisomerase that introduces negative supercoiling into DNA [59] thus facilitating both compaction and strand separation at biologically important DNA sequences. In this example, the energy level of DNA is raised. Other topoisomerases can reverse this effect, so relaxing DNA.

However, although topoisomerases can establish and maintain an equilibrium state of supercoiling, the processes of DNA replication and transcription generate transient changes in DNA supercoiling following the translocation of the protein complexes along the DNA [56]. These transients arise as a direct consequence of the double-helical structure of DNA. When proteins such as RNA polymerase move along DNA, they do not track linearly along the molecule but instead, by following one or other of the grooves, rotate along a helical path [60]. Many protein complexes are much more bulky than DNA and their freedom to rotate around the DNA may be constrained by molecular crowding creating viscous drag [61]. In some cases, the polymerizing enzymes may even be restrained in a fixed spatial position by secondary physical attachment to extensive structures, such as membranes [62–65]. Under these circumstances, provided the rotation of the DNA molecule is also constrained, torsional strain is generated such that the DNA is overwound downstream of the advancing enzyme and underwound upstream (Fig. 6). This principle, first proposed by Liu and Wang [56], plays a key role in the genetic organization of chromosomes.

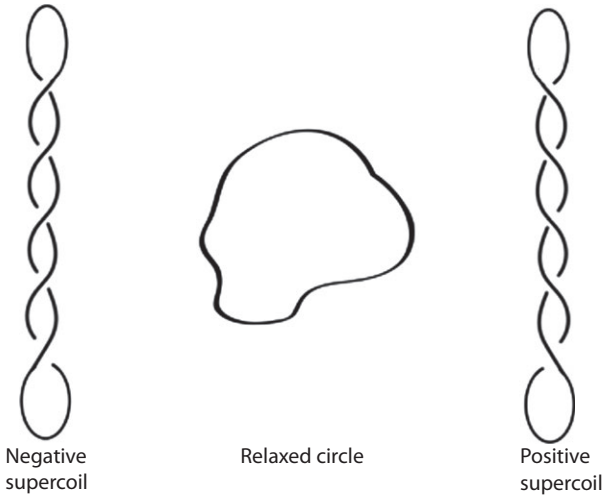

Fig. 5. Effect of superhelicity on the conformation of a small DNA circle. The figure shows the introduction of positive and negative cross-overs induced by positive and negative superhelicity, respectively.

Another important function of topoisomerases is to buffer DNA structure against temperature variations. Living organisms exist over a broad temperature range of approximately −15 °C to +120 °C and within this range, individual organisms can tolerate quite wide temperature variations. At the high end of the overall temperature range, the probability of adventitious melting, and consequently of errors in, for example, transcription initiation, is substantially increased. To counteract the possibility of such deleterious bubbles, extremophiles – for example, thermophilic bacteria and Archaea – often encode a reverse DNA gyrase that increases the twist of the DNA double helix. Indeed, recent calculations suggest such a strategy for stabilizing the double helix would enable the DNA of an extremophile, *Thermus thermophilus,* to retain sufficient stability for biological function up to a temperature of 106 °C, or 15 °C higher, than the maximum value for free DNA observed in the absence of topological constraints [66]. Even organisms that exist at more normal ambient conditions exert fine control over DNA structure to compensate for temperature changes. For example, the bacterium *Escherichia coli* maintains a constant superhelical stress over a temperature range of 17–37 °C [67]. This effect, presumably mediated by topoisomerases, compensates for temperature-dependent alteration of double-helical pitch over this temperature range.

In both the bacterial nucleoid and the eukaryotic nucleus, DNA is usually packaged as a negative supercoil [54] consistent with the preferential binding of negative supercoils by the most abundant nuclear (the nucleosome core particle and HMGB proteins) and bacterial nucleoid (HU, H-NS and FIS) proteins [33]. But why is DNA packaged in a higher energy state? Perhaps trivially, as in a ball of string, coiling is an efficient mode of packaging a long chain and so increases the compaction of long stretches of DNA. But the storage of negative supercoils also has the

17424658, 2015, 12, Downloaded from https://febs.onlinelibrary.wiley.com/doi/10.1111/febs.13307 by National Health Fund in Poland, Wiley Online Library on [01/08/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

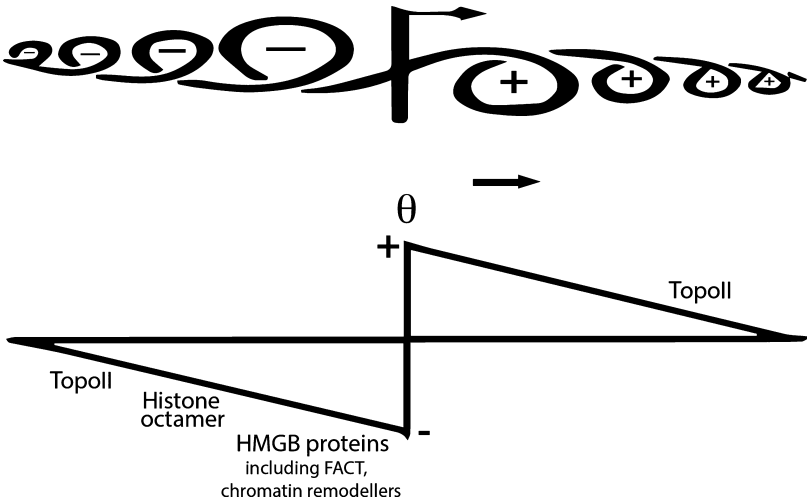

Fig. 6. Induction of superhelicity in a constrained DNA domain by a fixed elongating DNA translocase moving in the direction of the arrow. Negative superhelicity is generated upstream of the translocating enzyme and positive superhelicity downstream. (Upper) Gradients of superhelicity on each side of the translocase Adapted from Liu and Wang [56]. (Lower) Different architectural proteins might interact with the transient superhelicity. θ indicates superhelical density.

potential to facilitate the passage of DNA and RNA polymerases along a DNA template. If DNA is already packaged as a negative supercoil, release of these negative supercoils might effectively neutralize some or all of the positive superhelicity generated by an advancing enzyme and so, at least partially counteract any inhibitory effect of positive torsion on the procession of the melted bubble within the polymerase complex. After the passage of the polymerase, the negative superhelicity behind the enzyme would facilitate the repackaging of DNA [68]. Such a process would conserve negative superhelicity. The positive superhelicity in eukaryotes might be removed by relaxing topoisomerases, such as topoisomerase II, whereas in bacteria, DNA gyrase might use ATP to ultimately increase the average level of negative superhelicity. This distinction highlights a fundamental difference between bacteria and eukaryotes. By possessing DNA gyrase, whose activity is sensitive to ATP levels [69], bacteria – as well as blue–green algae, at least some mitochondria and chloroplasts [70–72] – can fine-tune the negative superhelicity of the genomic DNA so that its energy level reflects energy availability from external sources. Lacking DNA gyrase in the nucleus, this mechanism is not available to eukaryotes. Nevertheless, they can, and do, by protein binding conserve the negative superhelicity generated by DNA translocases. An increase in unconstrained (not protein-bound) negative superhelicity on chromatin decompaction [73] might arise both from the release of constrained supercoils or from the selective relaxation of positive superhelicity generated by transcription.

An under-appreciated aspect of the constraint of DNA superhelicity by DNA-binding proteins is that optimal superhelical density for binding is likely protein dependent. Although both the histone octamer and HMGB proteins constrain negative superhelicity, the octamer binds writhed DNA [34] and the HMGB proteins untwisted DNA [74]. Another protein, the bacteria nucleoid-associated protein HU constrains DNA that is likely both writhed and untwisted [75]. This difference is important, as the negative superhelicity generated by a DNA translocase is likely highest closest to the translocase and decays as the distance from the translocase increases [76]. In this context, it is intriguing that protein complexes containing an HMG domain, such as certain chromatin remodellers [77,78], and the FACT transcription elongation complex [79–81] are either translocases themselves, or act in close proximity to a translocation site [82]. Because untwisting is favoured by higher superhelical densities, it can be speculated that HMG domains stabilize the high nascent negative superhelicity generated by DNA translocases and only subsequently is some of this superhelicity constrained by the histone octamer.

An additional important facet of translocase function is the force applied to DNA by the enzyme complexes. For example, a transcribing bacterial RNA polymerase can exert a force of ~ 20 pN [83], whereas the maximal force exerted by bacteriophage Φ29 portal motor is > 100 pN [84]. Forces of this magnitude have the potential to alter the partition of superhelicity between twist and writhe [85]. Notably forces > 3 pN favour twist – both overtwisting and undertwisting – rather than writhe [85]. Depending on the distribution of the applied force this implies that, for example, bacterial RNA polymerase could facilitate transcription by uncoiling a negatively supercoiled DNA plectoneme downstream of the transcribing enzyme. Similarly a 4–5 pN force is required to extend and uncoil the 30 nm chromatin fibre [86]. Again, this is significantly less than the force exerted by RNA polymerase.

17424658, 2015, 12, Downloaded from https://febs.onlinelibrary.wiley.com/doi/10.1111/febs.13307 by National Health Fund in Poland, Wiley Online Library on [01/08/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

If packaged DNA constituted an energy store for modulating gene expression, the nature of the packaging might depend on energy availability. In bacteria, there is evidence that this is indeed the case. During the late stationary phase of growth when the cells are starved of energy, the nucleoid body collapses and the DNA is packaged by the highly abundant protein [87]. The proposed mode of packaging by *E. coli* Dps is very different from that of the abundant DNA binding proteins characteristic of exponential growth [88] and there is no evidence that it constrains DNA superhelicity, suggesting that high superhelical density is not a necessary concomitant of compaction. In eukaryotic nuclei, as in bacteria, loss of energy results in chromatin compaction [89]. Here, however, the molecular nature of the mechanisms involved is not yet understood. Nonetheless, it seems a reasonable proposition that the coupling between energy availability and DNA organization is tighter in bacteria than in eukaryotes.

Alternative DNA structures

Although a right-handed double helix is the canonical image of a DNA, it has long been recognized that the molecule can adopt a number of other biologically important structures. These include variations on the double helix such as bubbles, Z-DNA, cruciforms and slipped loops, three-stranded triple helices (H-DNA) and even distinct four-stranded structures, G-quadruplexes [21–25,90–93] (Fig. 7). H-DNA demonstrates a further structural capability of the DNA. The departure from the standard double-helical structure is accompanied by the formation of a different type of base-pair in H-DNA in which the third strand of the triple stranded complex forms Hoogsteen base pairs with a Watson–Crick paired double strand in the major groove of the latter [21,22]. Double-stranded sequences generating a G-quadruplex form a G-rich strand form another structure, the *i*-motif from the C-

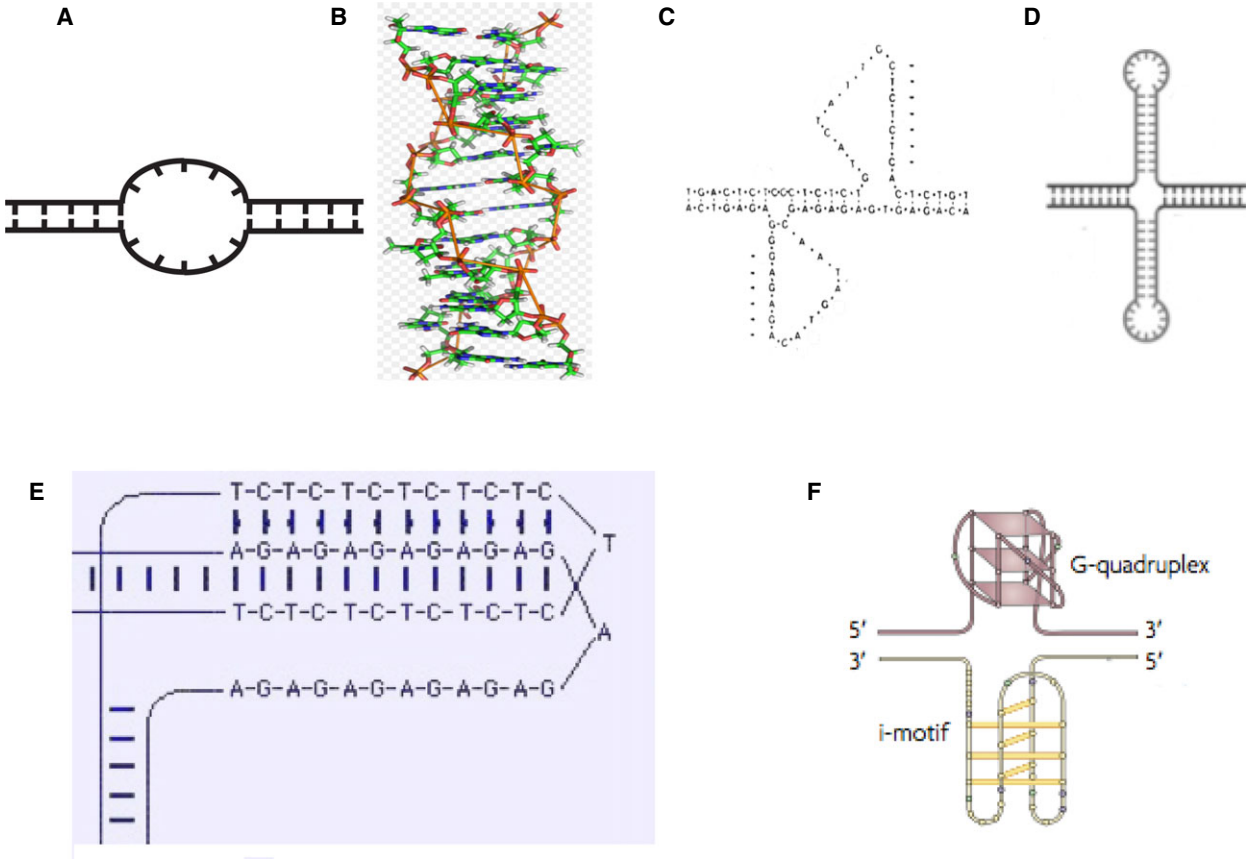

Fig. 7. Alternative DNA structures. Gallery of alternative structures showing (A) DNA bubble, (B) Z-DNA, (C) slipped loop, (D) cruciform, (E) H-DNA, (F) G-quadruplex/*i*-motif in double-stranded DNA. For the quadruplex/*i*-motif the structure assumed by the *i*-motif is likely pH dependent [19].

17424658, 2015, 12, Downloaded from https://febs.onlinelibrary.wiley.com/doi/10.1111/febs.13307 by National Health Fund in Poland, Wiley Online Library on [01/08/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

rich strand [25]. In the quadruplex, the four strands are connected by Hoogsteen base pairing [23,24], whereas the *i*-motif has yet another type of noncanonical pairing [25].

In addition to structures that can form over the normal range of physiological conditions, yet more different structures can be induced by the application of extensive and torsional forces greater than those normally encountered in the cell [94]. Such structures are very different to the classical B-type double helix [95–97] and include the underwound S-DNA with an estimated 33 bp per turn and the overwound P-DNA with 2.7 bp per turn [94,97], the latter possibly corresponding to the 'inside-out' structure originally proposed by Pauling and Corey [98]. The formation of S-DNA requires an extensive force of at least 60 pN, whereas P-DNA is only stable at a positive torque in excess of 30 $pN \cdot nm^{-1}$ [95].

The formation of alternative structures under physiological conditions is favoured not only by particular sequence organizations and base compositions, but also by the energetic environment of these DNA sequences. One of the simplest, and among the biologically most important, of these alternative structures is a DNA bubble in which strand separation occurs over a short sequence of base pairs generating a region consisting of two separated strands bounded by more stable double-helical stretches. Bubble formation is strongly sequence dependent, occurring most frequently at the TpA base-step [46] – the least stable of all 10 steps [47,48] (Table 1). Consequently, melting is favoured by agents – higher temperatures and negative superhelicity – that promote the unwinding of the double helix, and the distribution of preferred melting sites closely correlates with genetic function. For example, DNA sequences in close proximity to the transcription start point are, on average, enriched in the TpA base-step, as also are specific recombination sites [46].

Table 1. Melting energies of the 10 base-steps. Reproduced from Protozanova *et al.* [48].

Base step	Melting energy ($kcal \cdot mol^{-1}$)
TA	−0.12
TG/CA	−0.78
AA/TT	−1.04
AT	−1.27
AG/CT	−1.29
CG	−1.44
GA/TC	−1.66
GG/CC	−1.97
AC/GT	−2.04
GC	−2.70

Negative superhelicity also facilitates the formation of other alternative DNA structures, notably the left-handed Z-DNA and also cruciforms, as well as their slipped loop variants [91,92,99,100]. Although there has been debate about the occurrence of such structures *in vivo*, in some examples, sequences with the potential to form these structures are located in the vicinity of promoter regions [100,101]. G-Quadruplexes constitute another class of structure. They form from a G-rich single strand with a particular sequence organization and constitute the major structural motif at the single-stranded termini of eukaryotic telomeres [20,21]. Additionally, such sequences are frequently found in double-stranded form in internal positions, again often located close to promoter regions [102]. Again, like Z-DNA and slipped loops, their formation and that of the complementary C-rich hairpin is promoted by negative superhelicity [103–105].

The multiplicity of alternative DNA structures whose formation is dependent on the intrinsic torsional stress in the DNA begs the question of their function, if any. Their frequent association with promoter regions implies that they might facilitate, but not necessarily be essential for, transcription initiation. Although this process is presented as relatively simple in textbooks, in reality, the progression from polymerase binding to the escape of an actively transcribing polymerase presents conflicting topological problems (Fig. 8). After binding, the enzyme first mediates the melting of approximately slightly more than one turn of double-stranded DNA, thereby constraining negative superhelicity. But in a closed system this melting of one double-helical turn must be balanced by the generation of an equal and opposite positive superhelicity, which will be most manifest in the immediate vicinity of the polymerase. In the absence of abundant topoisomerases relaxing this positive superhelicity, it might be absorbed by alternative DNA structures converting them back to a simple double helix. However, there is another potential barrier to initiation – the escape of the polymerase. If the DNA is initially relaxed as the polymerase begins to move away from the promoter, the advancing complex will generate negative superhelicity behind and positive superhelicity in front. Absorption of the negative superhelicity by a DNA sequence with the potential to form an alternative structure might then enable the release of the polymerase complex from the promoter. One possible function of these structures is thus to act as a torsional buffer [104–108]. Similarly, in eukaryotic chromosomes, the positive superhelicity in front of polymerase might contribute to the facilitation of the unwrapping of DNA around nucleosomes [68].

17424658, 2015, 12, Downloaded from https://febs.onlinelibrary.wiley.com/doi/10.1111/febs.13307 by National Health Fund in Poland, Wiley Online Library on [01/08/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

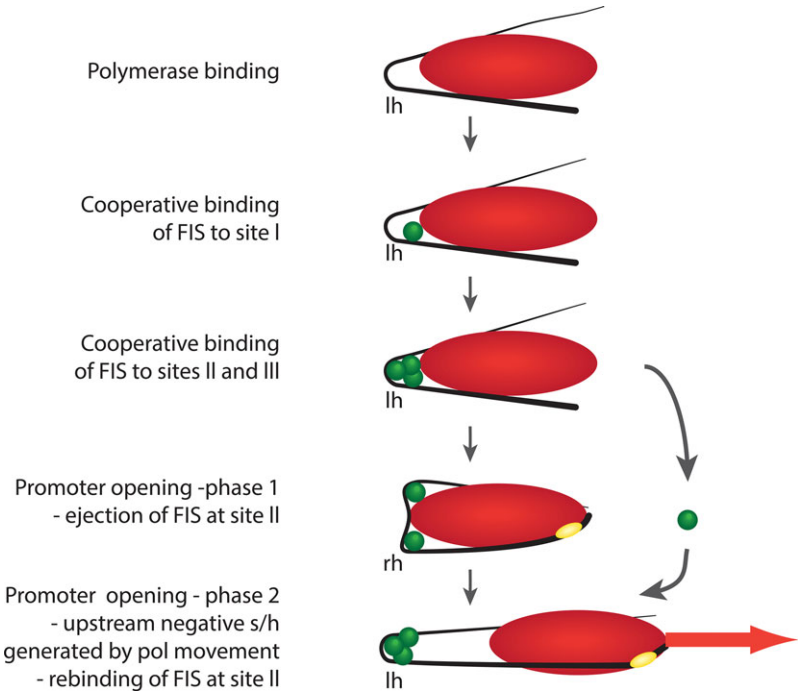

Fig. 8. A topological machine: topology of RNA polymerase initiation and escape. The enzyme (red ellipse) binds at a promoter site, melts the DNA over a limited region and then escapes from the promoter. These transitions are facilitated by a transcription factor binding upstream of and contacting the enzyme and are accompanied by changes in the path of the DNA around the polymerase and in the region upstream of the enzyme. Adapted from Muskhelishvili and Travers [108].

Although the ability of DNA to assume a variety of alternative structures is a highly visible manifestation of conformational flexibility, the molecule can undergo other more subtle biologically important transitions. One such is the property of coiling or writhing under torsional stress. Such coiling is favoured by increasing flexibility and hence higher A/T contents and, importantly, can be induced by both negative and positive superhelicity. Negative superhelicity also induces localized strand separation [46] and, in this case, any choice between writhing and strand separation will likely depend on the both the length and the organization of the sequence. In general, for a DNA stretch of low melting energy, the longer the sequence the greater the probability of responding to negative superhelicity by writhing, rather than by melting. This is because DNA superhelicity is not distributed uniformly along a DNA molecule but instead will be, on average, localized to those sequences that are most deformable. If a highly deformable sequence is short, and particularly if it is flanked by sequences for which deformation is energetically unfavourable, available superhelicity will be effectively concentrated in the short sequence resulting in a high local superhelical density [7,8,109]. By contrast, for a longer deformable sequence, the available superhelicity will be spread more extensively. Exemplars of short deformable sequences are the hexameric −10 region of strong bacterial promoters and similar sequences within budding yeast origins of DNA replication. In at least the former case, superhelicity favours strand separation at these sites [46]. By contrast, regions of low melting energy DNA extending to ~ 100 bp or more exist both upstream of strongly transcribed *E. coli* genes [110,111] and downstream of many budding yeast genes. In both cases, superhelicity introduced by DNA gyrase and/or by transcription is believed to induce DNA writhing [109]. The distinction between writhing and strand separation is crucial for biological function. Supercoiled DNA in *E. coli* has an average superhelical density of −0.05; that is, an average change of linking number from the fully relaxed form of −1/200 bp. Typically in the very active stable RNA promoters of *E. coli* the −10 hexamer is flanked by blocks of G/C-rich sequence of higher average melting temperature [109]. Such blocks potentially act as barriers and so can serve to localize the effects of torque – dependent on the intrinsic superhelicity or on direct enzymatic manipulation – to the short −10 region. The consequence is that the local superhelical density within this sequence is much higher than the average superhelical density and so melting is favoured [46]. This contrasts with the upstream regions of such promoters, which although also sensitive to supercoiling [111], are much more extensive and more uniform in base composition. Consequently, any superhelicity is distributed over more double-helical turns resulting in a lower local superhelical density than might be experi-

17424658, 2015, 12, Downloaded from https://febs.onlinelibrary.wiley.com/doi/10.1111/febs.13307 by National Health Fund in Poland, Wiley Online Library on [01/08/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

enced in a 'gated' −10 region, and thus for a preference for writhing rather than melting. The localization of linking number changes in supercoiled DNA is also an essential element in the formation of such structures as cruciforms and slipped loops [91,92].

The sequence-dependent bending anisotropy and elasticity of DNA have the potential to organize a DNA molecule in a preferred writhing configuration, or configurations, under superhelical stress. Such preferred configurations might contribute to the distinction of discrete domains within a single DNA molecule (e.g. (Fig. 9). When stabilized by proteins, such structures, which in an unconstrained DNA molecule are likely dynamic, might act as separate topological domains and enable differential modes of gene regulation within each domain. We suggest that effects such as these would be the expression of analogue information spanning several kilobases.

DNA and genetic organization

The physicochemical properties conferred by DNA sequence not only determine bending and melting preferences, but also correlate strongly with the genetic organization of both eukaryotic and bacterial chromosomes. In general, the coding sequences of genes have a G/C-rich bias [45,112]. In part, this is because the codons for the most abundant amino acids also have a G/C-rich bias [113,114]. The corollary is that noncoding DNA sequences, including introns as well as 5′ and 3′ flanking DNA sequences, are generally more A/T-rich. Indeed the most A/T-rich and most thermodynamically unstable DNA sequences in the *Saccharomyces cerevisiae* genome are located in 3′ flanking regions [110]. This distribution of base composition on a genomic scale implies that, on average, coding sequences are stiffer or less bendable, whereas noncoding sequences are both more flexible and more susceptible to strand separation. However, in apparent contradiction to these variations in flexibility, in eukaryotic chromosomes coding sequences have a higher nucleosome occupancy than noncoding sequences [45,112]. But, again, this pattern of occupancy is possibly related to another sequence-dependent physical property of the polymer, the higher intrinsic entropy of certain A/T-rich sequences [32].

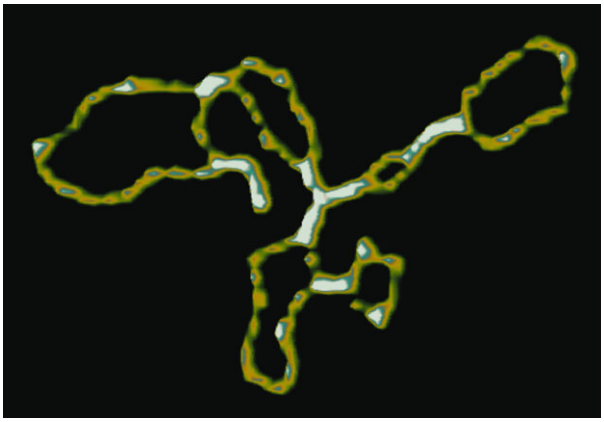

Fig. 9. Transient configuration of a negatively supercoiled mini chromosome. AFM visualization of a circular supercoiled pBR322 DNA molecule. In this configuration the molecule exhibits a distinct domain structure delimited by the duplex–duplex contacts in the centre of the picture. Previously unpublished image reproduced with kind permission from Sebastian Maurer.

The occurrence of the more A/T-rich sequences in the flanking regions of genes has functional significance. At the 5′-end of a transcription unit there is an obvious correlation with the requirement for RNA polymerase to melt DNA prior to transcription initiation. But at the 3′-ends of transcription units, polymerase dissociates and releases the constrained unwound DNA so that it reforms a double helix. One possibility is that such regions serve as topological sinks, absorbing by writhing any positive superhelicity generated in advance of the transcribing enzyme. This would block the transmission of any such superhelicity to a neighbouring gene with the potential for disrupting its chromatin structure. Instead, the writhed DNA would serve as an appropriate substrate for relaxation by topoisomerases; in particular, topoisomerase II, which is preferentially associated with actively transcribed genes [115]. Topoisomerase II, together with topoisomerase I, is also found in regions of low nucleosome occupancy at promoters [115]. However, measurement of the association of topoisomerase II with its optimal binding sites is precluded because of their highly repetitive and redundant nature.

The relationship of the physicochemical properties of DNA to chromosome organization and function in not only apparent at the level of individual genes and transcription units, but is also a feature of whole bacterial chromosomes. These chromosomes comprise, in general, a single circular DNA molecule which can vary in length from ~ 0.5 Mb to 6–10 Mb. Remarkably in these chromosomes, at least in most γ-Proteobacteria, gene order is highly conserved such that those genes that are highly expressed during exponential growth are clustered near the origin of DNA replication, whereas those that are more active during episodes of environmental stress resulting in the cessation of growth are more frequent in the vicinity of the replication termini [116,117]. However, not only is

17424658, 2015, 12, Downloaded from https://febs.onlinelibrary.wiley.com/doi/10.1111/febs.13307 by National Health Fund in Poland, Wiley Online Library on [01/08/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

there a gradient of gene organization from origin to terminus, but also this gradient correlates, on average, with a gradient of base composition so that in each replichore the most stable, G/C-rich, DNA is close to the origin while the least stable is at the terminus [117]. This average pattern of course includes wide variations at the level of individual genes. Yet another feature that exhibits a graded response from origin to terminus is the distribution of binding sites for DNA gyrase [116,118], a topoisomerase that inserts negative superhelical turns into DNA [55]. Again these are concentrated primarily in proximity to the origin of replication and thus create the potential for the DNA in this region to be more highly negatively supercoiled than that close to the terminus. This overall pattern of organization can couple chromosome structure to energy availability [69]. When bacteria are shifted to a fresh rich growth medium, ATP levels rise, activating DNA gyrase and thus increasing the negative superhelical density of the chromosome [60]. This would be localized to the origin-proximal region and would, in turn, activate the genes producing the necessary components for growth – the transcription and translation machinery – as well as providing an appropriate environment for DNA replication. Once DNA replication is initiated, the passage of the replisomes along the two replichores would by itself generate a gradient of superhelicity by the Liu/Wang principle [56], with the more negatively supercoiled DNA again being located closer to the origin and the more relaxed DNA close to the terminus. Again, by analogy to transcriptions, the DNA close to the terminus, in concert with topoisomerases, might act as a topological barrier between the two replichores. The bacterial chromosome thus functions as a topological machine in which the overall distribution of DNA sequences reflects the coupling between the processing of the replisomes and gene expression.

Although the Liu/Wang principle was initially conceived as applying to naked DNA, it is equally valid when considered in the context of higher order structures generated by DNA packaging. In eukaryotic nuclei, despite the existence of the 30 nm fibre *in vivo* being recently questioned [118], the left-handed coiling of the nucleosome stacks responds to torsional forces – such as those generated by transcription – by unwinding on application of positive torsion and correspondingly rewinding with applied negative torsion [119].

Informational capacity

Although the physicochemical properties of DNA determine its dynamic roles in the context of enzymatic manipulation, how are they integrated with the primary function of DNA as an information store? Like DNA, a fundamental property of RNA is to encode protein sequences, but unlike a DNA genome, an RNA genome must combine both information storage and its functional expression during the translation of the nucleotide sequence into protein. These two requirements are not necessarily wholly compatible. For example, when there are strong selective pressures to maintain the integrity of the genomic nucleotide sequence, the option of regulating translation by modulating the half-life of an RNA molecule is effectively excluded. Perhaps more tellingly, in known present-day biological systems, RNA molecules that function as messengers, and also as genomes (e.g. those of poliovirus and Qβ bacteriophage) are, relative to most genomic DNA molecules, very short, comprising only a few thousand nucleotides. An advantage of the separation of responsibilities between the two types of polynucleotide is that the juxtaposition and catenation of individual genes into much longer molecules enhances the potential regulatory repertoire of gene expression. In particular, the coordination of gene expression can be facilitated at the local level by the structural interplay arising from the transcriptional activity of adjacent genes, depending on whether they are organized in tandem or transcription is convergent or divergent. At a higher level of structural organization the continuity afforded by a single DNA double helix in a chromosome permits the organization of genes, and hence the available DNA information, into distinct structural and functional domains comprising many protein-coding elements. Apart from these considerations, the ability of DNA to accrete more and more packets of information – as genes or as regulatory elements – into longer and longer chromosomal molecules, is arguably an important factor contributing to increases in organismal complexity. Indeed, the postulated usurping of RNA by DNA as the informational store within a cell by itself epitomizes an evolutionary increase in biological complexity; one that enables increased possibilities for information storage – such as the differential encoding on complementary strands – and processing and hence, provides a substrate for further natural selection.

Any increase in the length of chromosomal DNA molecules should be coupled to – and possibly limited by – mechanisms for the generation and maintenance of genomic integrity. In this context, a relevant biological example is provided by ciliates – a group of single-celled organisms including the causal agents of malaria and sleeping sickness – in which regulation of DNA-directed

17424658, 2015, 12, Downloaded from https://febs.onlinelibrary.wiley.com/doi/10.1111/febs.13307 by National Health Fund in Poland, Wiley Online Library on [01/08/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

gene expression is separated from maintenance of the germ line. These organisms contain two types of nuclei. One, the micronucleus, containing the complete diploid genome, serves as the germ line and does not express genes [120], whereas the second, the polyploid macronucleus contains a highly edited and fragmented version of the genome, and serves as the vehicle controlling gene expression [121,122]. Only the micronucleus undergoes mitotic chromosomal segregation [123].

Conclusions

Although, quite rightly, the ability of DNA to code for protein sequences is often emphasized, equally important is the encoding of information that enables both the packaging of the polymer and the regulation of the expression and accessibility of the protein-coding information. This latter property depends to a much greater extent on the sequence-dependent physicochemical properties of the molecule and thus on the cumulative properties of a succession of a small number of base pairs. This analogue characteristic contrasts with the digital encoding of protein sequences. A further crucial aspect of DNA function is the dynamic nature of the structural transitions observed during its replication and transcription. These transitions involve both deviations from the canonical double helix and also topological transformations of the trajectory of the double helix facilitating both packaging and regulation. Taken together, these properties enable DNA to function as a supremely efficient and versatile coding device.

Author contributions

Both authors contributed sections to the manuscript.

References

1 Watson JD & Crick FH (1953) Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **171**, 737–738.
2 Avery OT, Macleod CM & McCarty M (1944) Studies on the chemical nature of the substance inducing transformation of pneumococcal types: induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III. *J Exp Med* **79**, 137–158.
3 Chargaff E, Lipshitz R, Green C & Hodes ME (1951) The composition of the deoxyribonucleic acid of salmon sperm. *J Biol Chem* **192**, 223–230.
4 deChadarevian S & Kamminga H (2002) Representations of the Double Helix. Whipple Museum of the History of Science, Cambridge, UK.
5 von Neumann J (1958) The Computer and the Brain. Yale University Press, New Haven, CT.
6 Marr C, Geertz M, Hütt MT & Muskhelishvili G (2008) Dissecting the logical types of network control in gene expression profiles. *BMC Syst Biol* **2**, 18.
7 Harris SA, Laughton CA & Liverpool TB (2008) Mapping the phase diagram of the writhe of DNA nanocircles using atomistic molecular dynamics simulations. *Nucleic Acids Res* **36**, 21–29.
8 Sayar M, Avşaroğlu B & Kabakçioğlu A (2010) Twist-writhe partitioning in a coarse-grained DNA minicircle model. *Phys Rev E Stat Nonlin Soft Matter Phys* **81**, 041916.
9 Arnott S, *et al.* (1966) Molecular and crystal structures of double- helical RNA. III. An 11-fold molecular model and comparison of the agreement between the observed and calculated three- dimensional diffraction data for 10 and 11-fold models. *J Mol Biol* **27**, 535–548.
10 Arnott S, Hutchinson F, Spencer M, Wilkins MH, Fuller W & Langridge R (1966) X-ray diffraction studies of double helical ribonucleic acid. *Nature* **211**, 227–232.
11 Dickerson RE, Drew HR, Conner BN, Wing RM, Fratini AV & Kopka ML (1982) The anatomy of A-, B-, and Z-DNA. *Science* **216**, 475–485.
12 Arnott S (2006) Historical article: DNA polymorphism and the early history of the double helix. *Trends Biochem Sci* **31**, 349–354.
13 Franklin RE & Gosling RG (1953) Molecular configuration in sodium thymonucleate. *Nature* **171**, 740–741.
14 Hormeño S, Ibarra B, Carrascosa JL, Valpuesta JM, Moreno-Herrero F & Arias-Gonzalez JR (2011) Mechanical properties of high-G.C content DNA with A-type base-stacking. *Biophys J* **100**, 1996–2005.
15 Bailly C, Waring MJ & Travers AA (1995) Effects of base substitutions on the binding of a DNA-bending protein. *J Mol Biol* **253**, 1–7.
16 Virstedt J, Berge T, Henderson RM, Waring MJ & Travers AA (2004) The exocyclic purine 2–amino group influences the stiffness of the DNA double helix. *J Struct Biol* **148**, 66–85.
17 Tseng YD, Ge H, Wang X, Edwardson JM, Waring MJ, Fitzgerald WJ & Henderson RM (2005) Atomic force microscopy study of the structural effects induced by echinomycin binding to DNA. *J Mol Biol* **345**, 745–758.
18 Hunter CA (1993) Sequence-dependent DNA structure. The role of base stacking interactions. *J Mol Biol* **230**, 1025–1054.
19 Gallego J, Luque FJ, Orozco M, Burgos C, Alvarez-Builla J, Rodrigo MM & Gago F (1994) DNA sequence-specific reading by echinomycin: role of

17424658, 2015, 12, Downloaded from https://febs.onlinelibrary.wiley.com/doi/10.1111/febs.13307 by National Health Fund in Poland, Wiley Online Library on [01/08/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

hydrogen bonding and stacking interactions. *J Med Chem* **7**, 1602–1609.
20 Hoogsteen K (1963) The crystal and molecular structure of a hydrogen-bonded complex between 1–methylthymine and 9–methyladenine. *Acta Crystallogr* **16**, 907–916.
21 Htun H & Dahlberg JE (1988) Single strands, triple strands, and kinks in H-DNA. *Science* **241**, 1791–1796.
22 Harvey SC, Luo J & Lavery R (1988) DNA stem-loop structures in oligopurine-oligopyrimidine triplexes. *Nucleic Acids Res* **16**, 11795–11809.
23 Sundquist WI & Klug A (1989) Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops. *Nature* **342**, 825–829.
24 Kang C, Zhang X, Ratliff R, Moyzis R & Rich A (1992) Crystal structure of four-stranded *Oxytricha* telomeric DNA. *Nature* **356**, 126–131.
25 Lyamichev VI, Mirkin SM & Frank-Kamenetskii MD (1985) A pH-dependent structural transition in the homopurine-homopyrimidine tract in superhelical DNA. *J Biomol Struct Dyn* **3**, 327–338.
26 Inoue S, Sugiyama S, Travers AA & Ohyama T (2007) Self-assembly of double-stranded DNA molecules at nanomolar concentrations. *Biochemistry* **46**, 164–171.
27 Baldwin GS, Brooks NJ, Robson RE, Wynveen A, Goldar A, Leikin S, Seddon JM & Kornyshev AA (2008) DNA double helices recognize mutual sequence homology in a protein free environment. *J Phys Chem B* **112**, 1060–1064.
28 Kornyshev AA & Leikin S (2001) Sequence recognition in the pairing of DNA duplexes. *Phys Rev Lett* **86**, 3666–3669.
29 Koo HS, Wu HM & Crothers DM (1986) DNA bending at adenine–thymine tracts. *Nature* **320**, 501–506.
30 Zhurkin VB (1983) Specific alignment of nucleosomes on DNA correlates with periodic distribution of purine-pyrimidine and pyrimidine-purine dimers. *FEBS Lett* **158**, 293–297.
31 Satchwell SC, Drew HR & Travers AA (1986) Sequence periodicities in chicken nucleosome core DNA. *J Mol Biol* **191**, 659–675.
32 Travers AA, Muskhelishvili G & Thompson JMT (2012) DNA information; from digital code to analogue structure. *Philos Trans R Soc Lond A* **370**, 2960–2986.
33 Luijsterburg MS, White MF, van Driel R & Dame RT (2008) The major architects of chromatin: architectural proteins in bacteria, archaea and eukaryotes. *Crit Rev Biochem Mol Biol* **43**, 393–418.
34 Luger K, Mäder AW, Richmond RK, Sargent DF & Richmond TJ (1997) Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**, 251–260.
35 Travers AA & Klug A (1987) The bending of DNA in nucleosomes and its wider implications. *Philos Trans R Soc Lond B* **317**, 537–561.
36 Gartenberg MR & Crothers DM (1988) DNA sequence determinants of CAP-induced bending and protein binding affinity. *Nature* **333**, 824–829.
37 Ntambi JM, Marini JC, Bangs JD, Hajduk SL, Jimenez HE, Kitchin PA, Klein VA, Ryan KA & Englund PT (1984) Presence of a bent helix in fragments of kinetoplast DNA minicircles from several trypanosomatid species. *Mol Biochem Parasitol* **12**, 273–286.
38 Diekmann S (1986) Sequence specificity of curved DNA. *FEBS Lett* **195**, 53–56.
39 Brukner I, Diakic M, Savic A, Susic S, Pongor S & Suck D (1993) Evidence for opposite groove-directed curvature of GGGCCC and AAAAAA tracts. *Nucleic Acids Res* **21**, 1025–1029.
40 Olson WK, Gorin AA, Lu XJ, Hock LM & Zhurkin VB (1998) DNA sequence-dependent deformability deduced from protein–DNA crystal complexes. *Proc Natl Acad Sci USA* **95**, 11163–11168.
41 Coll M, Frederick CA, Wang AH & Rich A (1987) A bifurcated hydrogen-bonded conformation in the d (A.T) base pairs of the DNA dodecamer d (CGCAAATTTGCG) and its complex with distamycin. *Proc Natl Acad Sci USA* **84**, 8385–8389.
42 Nelson HC, Finch JT, Luisi BF & Klug A (1987) The structure of an oligo(dA).oligo(dT) tract and its biological implications. *Nature* **330**, 221–226.
43 DiGabriele AD, Sanderson MR & Steitz TA (1989) A DNA dodecamer containing an adenine tract crystallizes in a unique lattice and exhibits a new bend. *Proc Natl Acad Sci USA* **86**, 1816–1820.
44 Ulanovsky L, Bodner M, Trifonov EN & Choder M (1986) Curved DNA: design, synthesis, and circularization. *Proc Natl Acad Sci USA* **83**, 862–866.
45 Struhl K & Segal E (2013) Determinants of nucleosome positioning. *Nat Struct Mol Biol* **20**, 267–273.
46 Drew HR, Weeks JR & Travers AA (1985) Negative supercoiling induces spontaneous unwinding of a bacterial promoter. *EMBO J* **4**, 1025–1032.
47 SantaLucia J Jr (1998) A unified view of polymer, dumbbell and oligonucleotide nearest-neighbor thermodynamics. *Proc Natl Acad Sci USA* **95**, 1460–1465.
48 Protozanova E, Yakovchuk P & Frank-Kamenetskii MD (2004) Stacked–unstacked equilibrium at the nick site of DNA. *J Mol Biol* **342**, 775–785.
49 Yan J & Marko J (2004) Localized single-stranded bubble mechanism for cyclization of short double helix DNA. *Phys Rev Lett* **93**, 108108.
50 Vinograd J, Lebowitz J, Radloff R, Watson R & Laipis P (1965) The twisted circular form of polyoma viral DNA. *Proc Natl Acad Sci USA* **53**, 1104–1111.

17424658, 2015, 12, Downloaded from https://febs.onlinelibrary.wiley.com/doi/10.1111/febs.13307 by National Health Fund in Poland, Wiley Online Library on [01/08/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

51 Cozzarelli NR, Boles TC & White JH (1990) Primer on the topology and geometry of DNA supercoiling. In DNA Topology and its Biological Effects (Cozzarelli NR & Wang JC, eds), pp. 139–184. Cold Spring Harbor Laboratory Press, NY.

52 Stonington OG & Pettijohn DE (1971) The folded genome of *Escherichia coli* isolated in a protein–DNA-RNA complex. *Proc Natl Acad Sci USA* **68**, 6–9.

53 Worcel A & Burgi E (1972) On the structure of the folded chromosome of *Escherichia coli*. *J Mol Biol* **71**, 127–147.

54 Bauer W (1978) Structure and reactions of closed duplex DNA. *Ann Rev Biophys Bioeng* **7**, 287–313.

55 Wang JC (2002) Cellular roles of DNA topoisomerases: a molecular perspective. *Nat Rev Mol Cell Biol* **3**, 430–440.

56 Liu LF & Wang JC (1987) Supercoiling of the DNA template during transcription. *Proc Natl Acad Sci USA* **84**, 7024–7027.

57 Wu HY, Shyy S, Wang JC & Liu LF (1988) Transcription generates positively and negatively supercoiled domains in the template. *Cell* **53**, 433–440.

58 Tsao YP, Wu HY & Liu LF (1989) Transcription-driven supercoiling of DNA: direct biochemical evidence from in vitro studies. *Cell* **56**, 111–118.

59 Gellert M, Mizuuchi K, O'Dea MH & Nash HA (1976) DNA work: an enzyme that introduces superhelical turns into DNA. *Proc Natl Acad Sci USA* **3**, 3872–3876.

60 Harada Y, Ohara O, Takatsuki A, Itoh H, Shimamoto N & Kinosita K Jr (2001) Direct observation of DNA rotation during transcription by *Escherichia coli* RNA polymerase. *Nature* **409**, 113–115.

61 Wang JC & Liu LF (1990) DNA replication: topological aspects and the roles of DNA topoisomerases. In DNA Topology and its Biological Effects (Cozzarelli NR & Wang JC, eds), pp. 321–340. Cold Spring Harbor Laboratory Press, New York.

62 Lodge J, Kazik T & Berg DE (1989) Formation of supercoiling domains in plasmid pBR322. *J Bacteriol* **171**, 2181–2187.

63 Lynch AS & Wang JC (1993) Anchoring of DNA to the bacterial cytoplasmic membrane through cotranscriptional synthesis of polypeptides encoding membrane proteins or proteins for export: a mechanism of plasmid hypernegative supercoiling in mutants deficient in DNA topoisomerase I. *J Bacteriol* **175**, 1645–1655.

64 Bowater RP, Chen D & Lilley DMJ (1994) Elevated unconstrained supercoiling of plasmid DNA generated by transcription and translation of the tetracycline resistance gene in eubacteria. *Biochemistry* **33**, 9266–9275.

65 Ma D, Cook DN, Pon NG & Hearst JE (1994) Efficient anchoring of RNA polymerase in *Escherichia coli* during coupled transcription-translation of genes encoding integral inner membrane polypeptides. *J Biol Chem* **269**, 15362–15370.

66 Meyer S, Jost D, Theodorakopoulos N, Peyrard M, Lavery R & Everaers R (2013) Temperature dependence of the DNA double helix at the nanoscale: structure, elasticity, and fluctuations. *Biophys J* **105**, 1904–1914.

67 Goldstein E & Drlica K (1984) Regulation of bacterial DNA supercoiling: plasmid linking numbers vary with growth temperature. *Proc Natl Acad Sci USA* **81**, 4046–4050.

68 Studitsky VM, Clark DJ & Felsenfeld G (1994) A histone octamer can step around a transcribing polymerase without leaving the template. *Cell* **76**, 371–382.

69 van Workum M, van Dooren SJ, Oldenburg N, Molenaar D, Jensen PR, Snoep JL & Westerhoff HV (1996) DNA supercoiling depends on the phosphorylation potential in *Escherichia coli*. *Mol Microbiol* **20**, 351–360.

70 Castora FJ, Vissering FF & Simpson M (1983) The effect of bacterial DNA gyrase inhibitors on DNA synthesis in mammalian mitochondria. *Biochim Biophys Acta* **740**, 417–427.

71 Itoh R, Takahashi H, Toda K, Kuroiwa H & Kuroiwa T (1997) DNA gyrase involvement in chloroplast-nucleoid division in *Cyanidioschyzon merolae*. *Eur J Cell Biol* **73**, 252–258.

72 Wall MK, Mitchenall LA & Maxwell A (2004) *Arabidopsis thaliana* DNA gyrase is targeted to chloroplasts and mitochondria. *Proc Natl Acad Sci USA* **101**, 7821–7826.

73 Naughton C, Avlonitis N, Corless S, Prendergast JG, Mati IK, Eijk PP, Cockroft SL, Bradley M, Ylstra B & Gilbert N (2013) Transcription forms and remodels supercoiling domains unfolding large-scale chromatin structures. *Nat Struct Mol Biol* **20**, 387–395.

74 Murphy FV 4th, Sweet RM & Churchill MEA (1999) The role of intercalating residues in chromosomal high-mobility-group protein DNA binding, bending and specificity. *EMBO J* **18**, 6610–6618.

75 Guo F & Adhya S (2007) Spiral structure of *Escherichia coli* HUαβ provides foundation for DNA supercoiling. *Proc Natl Acad Sci USA* **104**, 4309–4314.

76 Meyer S & Beslon G (2014) Torsion-mediated interaction between adjacent genes. *PLoS Comput Biol* **10**, e1003785.

77 Mohrmann L & Verrijzer CP (2005) Composition and functional specificity of SWI2/SNF2 class chromatin remodeling complexes. *Biochim Biophys Acta* **1681**, 59–73.

17424658, 2015, 12, Downloaded from https://febs.onlinelibrary.wiley.com/doi/10.1111/febs.13307 by National Health Fund in Poland, Wiley Online Library on [01/08/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

78 Bartholomew B (2014) Regulating the chromatin landscape: structural and mechanistic perspectives. *Annu Rev Biochem* **83**, 671–696.
79 Orphanides G, Wu WH, Lane WS, Hampsey M & Reinberg D (1999) The chromatin-specific transcription elongation factor FACT comprises human Spt16 and SSRP1 proteins. *Nature* **400**, 284–288.
80 Brewster NK, Johnston GC & Singer RA (2001) A bipartite yeast SSRP1 analog comprised of Pob3 and Nhp6 proteins modulates transcription. *Mol Cell Biol* **21**, 3491–3502.
81 Formosa T, Eriksson P, Wittmeyer J, Ginn J, Yu Y & Stillman DJ (2001) Spt16-Pob3 and the HMG protein Nhp6 combine to form the nucleosome-binding factor SPN. *EMBO J* **2**, 3506–3517.
82 Mason PB & Struhl K (2003) The FACT complex travels with elongating RNA polymerase II and is important for the fidelity of transcriptional initiation in vivo. *Mol Cell Biol* **23**, 8323–8333.
83 Yin H, Wang MD, Svoboda K, Landick R, Block SM & Gelles J (1995) Transcription against an applied force. *Science* **270**, 1653–1657.
84 Rickgauer JP, Fuller DN, Grimes S, Jardine PJ, Anderson DL & Smith DE (2008) Portal motor velocity and internal force resisting viral DNA packaging in bacteriophage Φ29. *Biophys J* **94**, 159–167.
85 Strick TR, Allemand JF, Bensimon D & Croquette V (1998) Behavior of supercoiled DNA. *Biophys J* **74**, 2016–2028.
86 Kruithof M, Chien FT, Routh A, Logie C, Rhodes D & van Noort J (2009) Single-molecule force spectroscopy reveals a highly compliant helical folding for the 30-nm chromatin fiber. *Nat Struct Mol Biol* **16**, 534–540.
87 Wolf SG, Frenkiel D, Arad T, Finkel SE, Kolter R & Minsky A (1999) DNA protection by stress-induced biocrystallization. *Nature* **400**, 83–85.
88 Grant RA, Filman DJ, Finkel SE, Kolter R & Hogle JM (1998) The crystal structure of Dps, a ferritin homolog that binds and protects DNA. *Nat Struct Biol* **5**, 294–303.
89 Visvanathan A, Ahmed K, Even-Faitelson L, Lleres D, Bazett-Jones DP & Lamond AI (2013) Modulation of higher order chromatin conformation in mammalian cell nuclei can be mediated by polyamines and divalent cations. *PLoS One* **8**, e67689.
90 Crawford JL, Kolpak FJ, Wang AH, Quigley GJ, van Boom JH, van der Marel G & Rich A (1980) The tetramer d(CpGpCpG) crystallizes as a left-handed double helix. *Proc Natl Acad Sci USA* **77**, 4016–4020.
91 Lilley DMJ (1980) The inverted repeat as a recognizable structural feature in supercoiled DNA molecules. *Proc Natl Acad Sci USA* **77**, 6468–6472.
92 Mace HAF, Pelham HRB & Travers AA (1983) Association of an S_1 nuclease-sensitive structure with short direct repeats 5′ of *Drosophila* heat shock genes. *Nature* **304**, 555–557.
93 Minyat EE, Khomyakova EB, Petrova MV, Zdobnov EM & Ivanov VI (1995) Experimental evidence for slipped loop DNA, a novel folding type for polynucleotide chain. *J Biomol Struct Dyn* **13**, 523–527.
94 Bustamante C, Bryant Z & Smith SE (2003) Ten years of tension: single molecule DNA mechanics. *Nature* **421**, 423–427.
95 Smith SB, Finzi L & Bustamante C (1992) Direct measurements of the elasticity of single DNA molecules using elastic beads. *Science* **258**, 1122–1126.
96 Strick TR, Allemand JF, Bensimon D, Lavery R & Croquette V (1996) The elasticity of a single supercoiled DNA molecule. *Science* **271**, 1835–1837.
97 Allemand JF, Bensimon D, Lavery R & Croquette V (1998) Stretched and overwound DNA forms a Pauling-like structure with exposed bases. *Proc Natl Acad Sci USA* **95**, 14152–14157.
98 Pauling L & Corey RB (1953) A proposed structure for nucleic acids. *Proc Natl Acad Sci USA* **39**, 84–97.
99 Singleton CK, Klysik J, Stirdivant SM & Wells RD (1982) Left-handed Z-DNA is induced by supercoiling in physiological ionic conditions. *Nature* **299**, 312–316.
100 Glaser RL, Thomas GH, Siegfried E, Elgin SC & Lis JT (1990) Optimal heat-induced expression of the Drosophila hsp26 gene requires a promoter sequence containing (CT)n. (GA)n repeats. *J Mol Biol* **211**, 751–761.
101 Schroth GP, Chou PJ & Ho PS (1992) Mapping Z-DNA in the human genome. Computer-aided mapping reveals a nonrandom distribution of potential Z–DNA-forming sequences in human genes. *J Biol Chem* **267**, 11846–11855.
102 Huppert JL & Balasubramanian S (2007) G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res* **35**, 406–413.
103 Voloshin ON, Veselkov AG, Belotserkovskii BP, Danilevskaya ON, Pavlova MN, Dobrynin VN & Frank-Kamenetskii MD (1992) An eclectic DNA structure adopted by human telomeric sequence under superhelical stress and low pH. *J Biomol Struct Dyn* **9**, 643–652.
104 Kouzine F, Sanford S, Elisha-Feil Z & Levens D (2008) The functional response of upstream DNA to dynamic supercoiling. *Nat Struct Mol Biol* **15**, 146–154.
105 Sun D & Hurley LH (2009) The importance of negative superhelicity in inducing the formation of G-quadruplex and i-motif structures in the c-Myc promoter: implications for drug targeting and control of gene expression. *J Med Chem* **52**, 2863–2874.

17424658, 2015, 12, Downloaded from https://febs.onlinelibrary.wiley.com/doi/10.1111/febs.13307 by National Health Fund in Poland, Wiley Online Library on [01/08/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

106 Carnevali F, Caserta M & Di Mauro E (1984) Transitions in topological organization of supercoiled DNA domains as a potential regulatory mechanism. *J Biol Chem* **259**, 12633–12643.
107 Kouzine F, Gupta A, Baranello L, Wojtowicz D, Ben-Aissa K, Liu J, Przytycka TM & Levens D (2013) Transcription-dependent dynamic supercoiling is a short-range genomic force. *Nat Struct Mol Biol* **20**, 396–403.
108 Muskhelishvili G & Travers A (2013) Integration of syntactic and semantic properties of the DNA code reveals chromosomes as thermodynamic machines converting energy into information. *Cell Mol Life Sci* **70**, 4555–4567.
109 Muskhelishvili G & Travers A (2013) DNA structure and bacterial nucleoid-associated proteins. In: Bacterial Gene Regulation and Transcriptional Networks, Chapter 2 (Babu MM, ed), pp. 37–52. Horizon Scientific Press, Norwich, UK.
110 Travers AA & Muskhelishvili G (2013) DNA thermodynamics shape chromosome organisation and topology. *Biochem Soc Trans* **41**, 548–553.
111 Opel ML, Aeling KA, Holmes WM, Johnson RC, Benham CJ & Hatfield GW (2004) Activation of transcription initiation from a stable RNA promoter by a Fis protein-mediated DNA structural transmission mechanism. *Mol Microbiol* **53**, 665–674.
112 Mavrich TN, Ioshikhes IP, Venters BJ, Jiang C, Tomsho LP, Qi J, Schuster SC, Albert I & Pugh BF (2008) A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome. *Genome Res* **18**, 1073–1083.
113 Crick FHC (1968) The origin of the genetic code. *J Mol Biol* **38**, 367–379.
114 Travers A (2006) The evolution of the genetic code revisited. *Orig Life Evol Biosph* **36**, 549–555.
115 Sperling AS, Jeong KS, Kitada T & Grunstein M (2011) Topoisomerase II binds nucleosome-free DNA and acts redundantly with topoisomerase I to enhance recruitment of RNA Pol II in budding yeast. *Proc Natl Acad Sci USA* **108**, 12693–12698.
116 Sobetzko P, Travers A & Muskhelishvili G (2012) Gene order and chromosome dynamics coordinate spatiotemporal gene expression during the bacterial growth cycle. *Proc Natl Acad Sci USA* **109**, E49–E50.
117 Sobetzko P, Glinkowska M, Travers A & Muskhelishvili G (2013) DNA thermodynamic stability and supercoil dynamics determine the gene expression program during the bacterial growth cycle. *Mol BioSyst* **9**, 1643–1651.
118 Jeong KS, Ahn J & Khodursky AB (2004) Spatial patterns of transcriptional activity in the chromosome of *Escherichia coli*. *Genome Biol* **5**, R86.
119 Maeshima K, Hihara S & Eltsov M (2010) Chromatin structure; does the 30-nm fibre exist in vivo? *Curr Opin Cell Biol* **22**, 291–297.
120 Recouvreux P, Lavelle C, Barbi M, Condé e Silva N, Le Cam E, Victor JM & Viovy JL (2011) Linker histones incorporation maintains chromatin fiber plasticity. *Biophys J* **100**, 2726–2735.
121 Davidson L & LaFountain JR Jr (1975) Mitosis and early meiosis in *Tetrahymena pyriformis* and the evolution of mitosis in the phylum Ciliophora. *Biosystems* **7**, 326–336.
122 Woodard J, Kaneshiro E & Gorovsky MA (1972) Cytochemical studies on the problem of macronuclear subnuclei in *Tetrahymena*. *Genetics* **70**, 251–260.
123 Yao MC, Choi J, Yokoyama S, Austerberry CF & Yao CH (1984) DNA elimination in *Tetrahymena*: a developmental process involving extensive breakage and rejoining of DNA at defined sites. *Cell* **36**, 433–440.
124 Johnson RE, Prakash L & Prakash S (2005) Biochemical evidence for the requirement of Hoogsteen base pairing for replication by human DNA polymerase iota. *Proc Natl Acad Sci USA* **102**, 10466–10471.
125 Hunter CA (1996) Sequence-dependent DNA structure. *BioEssays* **18**, 157–162.

17424658, 2015, 12, Downloaded from https://febs.onlinelibrary.wiley.com/doi/10.1111/febs.13307 by National Health Fund in Poland, Wiley Online Library on [01/08/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License